Earth Science Data Systems
Reuse Working Group Case Study:
Simple, Scalable, Script-based Science Processor
for Measurements
(S4PM)

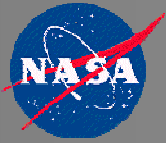Christopher Lynnes (Christopher.S.Lynnes@nasa.gov)

Goddard Earth Sciences Data and Information Services Center

NASA/GSFC

http://disc.gsfc.nasa.gov/techlab/s4pm/

4/29/2005
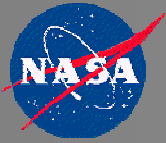
- Goal: highly automated processing of science data
- Background:
    – Originally developed so GES DAAC could scale up for MODIS Reprocessing (2000)
    – Then MODIS forward processing (2001)
    – Then AIRS processing (2002)
    – Then replacement of EOSDIS Core System processing s/w (2002)
    – Then On-Demand subsetting (2004)
    – Then Data Mining (2005)
- Software Aspects:
    – Scope: all automated science processing at GES DAAC
   – Code: ~26,000 lines of Perl

- Simple, Scalable, Script-based Science Processor (S4P) implements a data-driven automated processing system with an assembly-line metaphor

  - Work is executed at **stations** (Unix directories) based on text **work orders**
  - Output work orders are sent to downstream stations

work order → (Station 1) work order → (Station 2) work order → (Station 3) work order →
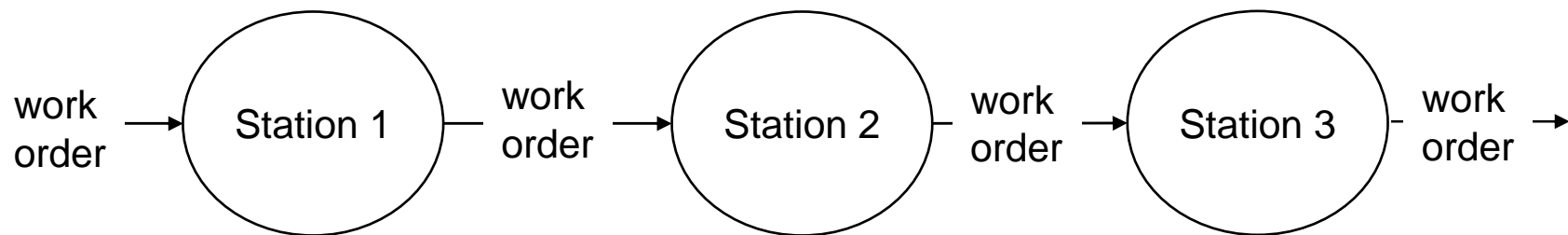
- Can be seen as a straightforward implementation of a data flow diagram
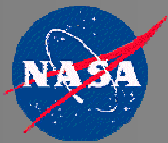
- Simple, Scalable, Script-based Science Processor
  - Simple: relatively small code base, simple GUIs
  - Scalable:
    - Up to all of the GES DAAC processing
    - Down to simple, single-algorithm scenarios
  - Script-Based: entirely in Perl
  - Science
    - Originally designed for science processing
    - But now used for automating other tasks
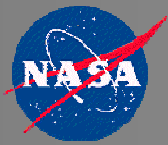
- ## Software Assets Reused
  - Reused Metaphor: assembly line directories from AVHRR 1-KM processing system at EROS Data Center
- ## Reusable Software Assets Created
  - S4P "kernel" and graphical monitor
    - Allows developers to implement their own processing or other automated system
    - University of South Florida: Direct Broadcast processing on clusters
    - LaRC: Calipso processing
    - Internal:
      - WHOM Ordering Engine: fault-tolerant submission of data orders
      - Watchtower: fault monitoring system for multiple machines/conditions
      - S4PA: Disk based data archive
  - S4PM
    - LaRC: MISR processing
    - EROS Data Center: ASTER On-Demand processing
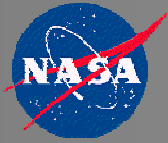
- **Benefits Realized**
  - Saved > $2M through replacement of predecessor system
    - Operations + COTS + sustaining engineering
  - Met schedule easily for each launch or campaign
  - Reliability and flexibility far exceed predecessor
- **Issues Encountered**
  - Too flexible? Code base has gotten a little messy over time…
- **Importance to meeting project goals or coping with constraints**
  - Key functions (e.g., data mining) could not have been realized without S4PM
  - Data processing rates could not have been realized without S4PM
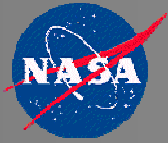
- Using Reusable Software Assets
  - Ran across AVHRR 1KM concept almost by chance in a site visit
- Building Reusable Assets
  - Building for internal reuse enabled external reuse
    - Keep it simple / standalone / independent of COTS
    - Modularity and loose coupling
  - Workshop held for S4PM developers at EDC and LaRC
  - Difficult to get software released outside of NASA
    - Patent release, ITAR, export control, security, copyright…
    - S4P release took 1 year; S4PM took 8 months
- Recommendations
  - Streamline NASA software release process
  - Use something like Tech. Readiness Levels to classify reusability